



**Lidil**

Revue de linguistique et de didactique des langues

**54 | 2016**

**La phrase en production d'écrits, approches nouvelles  
en didactique**

---

## La phrase : de la maîtrise de ses conventions graphiques à sa pertinence cognitive

*From the Control of the Graphic Conventions of the Sentence to Its Cognitive  
Relevance: An On-Line Study*

Florence Chenu et Émilie Ailhaud

---



### Édition électronique

URL : <http://journals.openedition.org/lidil/4024>

DOI : 10.4000/lidil.4024

ISSN : 1960-6052

### Éditeur

UGA Éditions/Université Grenoble Alpes

### Édition imprimée

Date de publication : 20 novembre 2016

Pagination : 15-34

ISBN : 978-2-84310-339-1

ISSN : 1146-6480

### Référence électronique

Florence Chenu et Émilie Ailhaud, « La phrase : de la maîtrise de ses conventions graphiques à sa pertinence cognitive », *Lidil* [En ligne], 54 | 2016, mis en ligne le 01 janvier 2017, consulté le 29 octobre 2020. URL : <http://journals.openedition.org/lidil/4024> ; DOI : <https://doi.org/10.4000/lidil.4024>

---

# La phrase : de la maîtrise de ses conventions graphiques à sa pertinence cognitive

*Florence Chenu et Émilie Ailhaud\**

## RÉSUMÉ

L'enseignement de la ponctuation, étroitement lié à la compréhension de la notion de « phrase », est un aspect délicat de la didactique de l'écrit. En effet, les conventions graphiques ne sont pas encore maîtrisées à l'entrée en collège, et du point de vue de la segmentation du texte, la notion de phrase syntaxique permet de mieux rendre compte des écrits des élèves. Cependant, malgré les difficultés typographiques, il est possible que la phrase, syntaxique ou graphique, soit bien l'unité de planification lors de l'écriture d'un texte. La question se pose alors du rapport entre la maîtrise de l'ensemble des conventions liées à la phrase et la pertinence de celle-ci d'un point de vue cognitif. Nous avons étudié ces deux aspects d'un point de vue développemental : les textes analysés sont des textes expositifs et narratifs composés par 40 élèves de CM2, 40 élèves de 5<sup>e</sup> et 40 élèves de 3<sup>e</sup>. Les erreurs liées aux marques de ponctuation diminuent avec le niveau scolaire. Les unités de planification évoluent également avec le niveau scolaire : les élèves de CM2 font plutôt des pauses longues en début de clauses mais des pauses courtes en début de phrase graphique, alors que les élèves de 5<sup>e</sup> et de 3<sup>e</sup> font essentiellement des pauses longues en début de phrase syntaxique. La planification « clause par clause », typique des CM2, est plus particulièrement présente chez les élèves utilisant les marques de ponctuation de façon non standard. L'utilisation de la phrase syntaxique lors de l'enseignement nous semble donc particulièrement intéressante à deux égards : d'une part car sa définition est claire, d'autre part car elle est pertinente d'un point de vue cognitif, puisqu'elle devient une unité de planification. Les connecteurs et marques de ponctuation dans leur diversité peuvent ensuite être discutés afin de trouver la meilleure marque de liaison entre deux phrases syntaxiques, la notion de phrase graphique émergeant par la suite.

---

\* Université Lyon 2, Laboratoire Dynamique du langage.

### ABSTRACT

*The teaching of punctuation, which is closely related to the understanding of the concept of “sentence” is a tricky aspect of the teaching of writing. Indeed, the graphic conventions are not mastered at the beginning of junior high school, and regarding the segmentation of a text, the notion of syntactic sentence can better reflect students’ writing. Despite typographical difficulties, it is possible that the sentence, with its syntactical or graphical definition, might well be the planning unit when writing texts. The question then arises of the relationships between the mastery of conventional markings of sentences and their relevance at the cognitive level. We studied these two aspects in a developmental perspective: we gathered and analyzed a narrative and an expository text from 40 5th grade students, 40 7th graders and 40 9th graders. Errors in punctuation marks decrease with grade level. Planning units are also changing with grade level: 5th graders use long pauses at the beginning of clauses but short pauses at the beginning of graphic sentence, while 7th and 9th graders essentially show long pauses at the beginning of syntactic sentence. “Clause by clause” planning, typical of 5th graders, is particularly present in students using punctuation marks in a nonstandard way. The use of the notion of syntactic sentence when teaching therefore seems particularly interesting for two reasons: first, because its definition is clearer than that of sentence, and second because it is relevant from a cognitive point of view, as it becomes a planning unit through grade level. Connectors and punctuation marks in their diversity can then be discussed in order to find the best connection between two syntactic sentences, the concept of graphic sentence could be brought afterwards.*

## 1. Introduction

La maîtrise d’un répertoire linguistique, qui englobe un large éventail de registres et de genres, ainsi que la capacité de s’adapter aux contextes communicationnels (Berman, 2006 ; Berman & Ravid, 2009 ; Ravid & Tolchinsky, 2002 ; Tolchinsky, 2006) se développe avec les années de scolarisation et avec les expériences linguistiques de l’individu et implique dans nos sociétés tant l’oral que l’écrit. Parmi les différences importantes qui caractérisent l’écrit par rapport à l’oral se trouve la question de la segmentation. L’écrit présente en effet de nombreux niveaux de segmentation visibles (caractères, mots, portions de phrases, phrases, paragraphes, etc.) dont certains seulement ont une contrepartie orale qui n’est d’ailleurs pas systématique. Cette segmentation, si l’élève y est exposé face à tout écrit, ne reçoit pas dans son enseignement le même détail d’explicitation à tous les niveaux. En particulier, la façon dont

les clauses<sup>1</sup> sont assemblées ou séparées pour former des unités (par l'usage de connecteurs et de la ponctuation) ne semble pas chose aisée du côté des scripteurs débutants aussi bien que du côté de l'enseignement. Un aspect délicat de la didactique de l'écrit concerne notamment l'enseignement de la ponctuation, dont la plurifonctionnalité est parfois peu identifiée par les enseignants (Dufour, 2014). La maîtrise de la ponctuation est liée étroitement à la compréhension de la notion de « phrase », elle-même difficile à définir (Béguelin, 2002 ; Berrendonner, 2002 ; Kleiber, 2003) et donc difficile à enseigner. Or, si la ponctuation est souvent vue comme une aide pour le lecteur, et est donc parfois rajoutée à posteriori lors de la relecture du texte, d'un point de vue cognitif, elle « constitue le lieu privilégié d'accomplissement de l'activité de planification » (Passerault, 1991, p. 91). Les frontières d'unités syntaxiques méritent donc d'une part d'être analysées par rapport aux normes de marquage typographique, et d'autre part d'être abordées du point de vue de leur fonction cognitive.

Dans un premier temps, nous interrogeons la question de la segmentation du texte et envisageons l'apport des données chronométriques dans la compréhension des processus rédactionnels. Nous nous proposons ensuite de mettre en lien les données chronométriques recueillies dans des textes d'élèves et le rapport à la norme que présentent les marques de ponctuation de ces textes.

## 2. Segmentation et dynamique de l'écriture

La terminologie concernant la manière de segmenter le discours varie d'une école théorique à l'autre. Il convient donc de définir précisément à quoi renvoient les termes que nous allons utiliser. La phrase est souvent considérée comme le domaine ultime de la syntaxe (Le Goffic, 1993, p. 8). Cependant la définition de la phrase se heurte au problème de la frontière entre le syntaxique et le textuel (Roulet, 2002). En effet, au critère de maximalité syntaxique s'ajoutent des critères de complétude sémantique et des critères de démarcation prosodiques ou typographiques (Kleiber, 2003). La question de l'autonomie est problématique

---

1. Nous utiliserons dans notre étude le terme de « clause » dans le sens proposé par Berman & Slobin (1994), qui correspond globalement au terme de « proposition » souvent employé dans les grammaires du français, telle la *Grammaire méthodique du français* (Riegel, Pellat & Rioul, 2002).

puisque, d'un point de vue sémantique tout au moins, toute phrase est en partie dépendante du contexte (Creissels, 2006, p. 12). D'un point de vue opératoire, la clause, telle que définie dans Berman et Slobin (1994) comme une unité comportant un seul prédicat unifié<sup>2</sup>, présente l'avantage d'être une unité minimale à partir de laquelle peuvent s'envisager des assemblages. Ces assemblages relèvent de mécanismes de subordination ou de coordination. Concernant les subordonnées, Creissels (1995, p. 324) en propose une révision du classement, en privilégiant le point de vue syntaxique, considérant que « le propre des structures phrastiques subordonnées [est] d'être assimilables à un constituant de la structure matrice ». Ceci revient à identifier, comme le propose Hunt (1970, p. 4), l'unité terminable minimale « *minimal terminable unit* », qui englobe les clauses entretenant des relations de dépendance syntaxique (c'est-à-dire une clause matrice avec ses subordonnées). Cette unité correspond à ce que nous avons nommé « phrase syntaxique » (PS). Ces PS peuvent parfois être coordonnées entre elles ou simplement juxtaposées, formant des unités supérieures typographiquement démarquées (majuscules initiales associées à des points)<sup>3</sup>. Il est donc possible d'identifier des « phrases graphiques » (PG), définies comme débutant par une majuscule et se terminant par un point, ce qui se rapproche le plus de la notion de « phrase » telle qu'elle est utilisée dans le langage courant ou dans l'enseignement. Ces PG reflètent ainsi en partie la façon dont le locuteur scripteur appréhende les phrases, l'enseignement de la notion de phrase mettant un accent particulier sur ses marques typographiques<sup>4</sup>.

Fayol (1989) propose de replacer la question de la segmentation dans le cadre de la linéarisation : le scripteur étant contraint de linéariser une organisation non nécessairement linéaire, il est amené à indiquer par des moyens linguistiques (connecteurs, ponctuation) les liens entre les clauses. C'est par cette contrainte que s'organisent des paquets de

---

2. Même s'il s'agit d'une unité relationnelle s'organisant autour d'un prédicat, il ne s'agit pas exactement de la « clause » au sens du modèle de Fribourg (Béguelin, 2002), puisqu'il n'est notamment pas fait référence ici à la notion d'acte énonciatif.

3. Pour un examen critique des notions traditionnelles de coordination et subordination, voir Creissels (1995, p. 307-311).

4. Le *Bulletin officiel* hors-série n° 3 du 19 juin 2008 suggère qu'en CP les élèves sachent « identifier les phrases d'un texte en s'appuyant sur la ponctuation (point et majuscule) ».

clauses qui répondent à la notion de phrase (graphique ou syntaxique). Selon Fayol (1989, p. 24), « la plus ou moins grande «force» des relations entretenues par les événements ou états décrits dans deux clauses successives sera le principal déterminant de la sélection de telle ou telle marque de ponctuation par l’auteur du texte ».

Nos travaux prennent en compte la dynamique de l’écriture en recueillant des données chronométriques au cours de la production écrite de texte (pauses, débits, etc.). Nos études consacrées aux caractéristiques lexicales et infra-lexicales des mots d’un texte en relation avec les données chronométriques ont permis de mettre au jour des influences immédiates, différées et anticipées dans la dynamique de l’écriture d’un mot dans un texte (Maggio, Lété, Chenu, Jisa & Fayol, 2012). Lors de ce travail, nous avons écarté une partie des mots du texte — ceux qui étaient précédés d’un signe de ponctuation ou d’un retour à la ligne — parce qu’ils ne nous semblaient pas comparables aux mots précédés de mots. Ceci nous a amenés à nous interroger sur le lieu de la planification dans nos textes. Certes, la segmentation en phrases graphiques avec une ponctuation conforme à la norme n’est pas encore bien maîtrisée en 6<sup>e</sup> (Fauchart, 2001 ; Paolacci & Favart, 2001). Cependant, la phrase s’est révélé une unité pertinente pour expliquer la dynamique de l’écriture pour nos scripteurs les plus avancés (Ailhaud, Chenu & Jisa, 2016). En effet, l’analyse des données chronométriques (pauses et débits) de ces textes a montré qu’avec l’avancée dans la scolarité, les scripteurs semblent planifier leur discours non plus clause par clause, mais à un niveau supérieur, comme la phrase syntaxique. Cette dernière semble alors devenir une unité réellement pertinente du point de vue cognitif lors de la planification du texte, ce qui rapproche le comportement des élèves de 3<sup>e</sup> de celui observé chez des adultes. La question se pose alors du rapport entre la maîtrise de l’ensemble des conventions liées à la phrase graphique canonique et la pertinence de celle-ci d’un point de vue cognitif.

### 3. Méthodologie

#### 3.1. *Participants*

Les données ont été collectées en France dans des établissements de zones urbaines de la région lyonnaise, dans trois écoles primaires et un collège ZEP/REP. La collecte de données a été entreprise en 2006 dans le cadre d’un projet financé par l’ANR (ReFlex ; resp. H. Jisa) et comprenait 3 groupes d’âge (10-11 ans – CM2, 12-13 ans – 5<sup>e</sup> et 14-15 ans –

3°), avec 40 participants par niveau scolaire. En tout, 120 élèves ont donc participé à cette étude.

### **3.2. Collecte des données**

Nous avons sollicité de la part de chaque individu la production de deux types de textes différents (narratif et expositif) à l'oral et à l'écrit. Une thématique commune a été sélectionnée pour les deux types de textes : celle du conflit entre les gens, qui s'était trouvée être porteuse lors d'études pilotes (Gayraud, 2000 ; Viguié-Simon, 2001). La thématique attendue des participants a été présentée au moyen d'une vidéo montrant une série de saynètes portant sur des problèmes entre les gens tels que la tricherie, les disputes, l'exclusion, etc. Ici seuls les textes produits à l'écrit sont pris en compte ; notre corpus est donc constitué de 240 textes (3 niveaux x 40 participants x (1 texte expositif + 1 texte narratif)). Ces textes ont été recueillis sur tablettes graphiques et par l'intermédiaire du logiciel *Eye and Pen* (Chesnet & Alamargot, 2005), dispositif qui permet de disposer sur la tablette un papier à grands carreaux sur lequel les sujets écrivent avec un stylo bille *Wacom intuos inking pen*.

### **3.3. Codage des données**

Trois types d'information ont été codés : la segmentation du texte, les écarts par rapport à la norme concernant la ponctuation et la longueur des pauses aux frontières d'unités syntaxiques. En annexe, un exemple de texte illustre ces codages.

#### **3.3.1. Segmentation du texte**

Chaque texte a été découpé selon plusieurs unités, telles que définies plus haut : les phrases graphiques (PG), débutant par une majuscule et se terminant par un point<sup>5</sup> ; les phrases syntaxiques (PS), qui sont des paquets de clauses en relation de dépendance syntaxique ; les clauses (C), au sens de Berman et Slobin (1994), unités comportant un seul prédicat unifié<sup>6</sup>.

---

5. Dans notre corpus, il arrivait qu'une des deux marques soit absente : une nouvelle PG était tout de même comptée dès lors qu'il y avait un point ou une majuscule.

6. Pour une description détaillée de la segmentation en clauses pour des textes en français, voir Gayraud (2000).

L'énoncé (1) correspond à une PG, composée elle-même de deux PS, séparées par «//»; la deuxième PS est composée de deux C, séparées par «/».

(1) *À deux heures trente tout le monde repartit en cours// et on apprend tous/ que la principale appellerait nos parents*<sup>7</sup>. [S3, 3<sup>e</sup>, narratif]<sup>8</sup>

Pour la suite de notre étude, nous utiliserons les termes PG, PS et C de manière mutuellement exclusive : les « frontières de C » correspondent à un début de nouvelle C qui ne soit pas une nouvelle PS ; les « frontières de PS » correspondent à un début de nouvelle PS qui ne soit pas une nouvelle PG. Cela est illustré dans l'exemple 2, qui reprend l'énoncé de l'exemple 1.

(2) [PG] *À deux heures trente tout le monde repartit en cours* [PS] *et on apprend tous* [C] *que la principale appellerait nos parents*. [S3, 3<sup>e</sup>, narratif]

Pour chaque texte, nous avons relevé la proportion de PG contenant 1 ou 2 PS : cela correspond à des PG plutôt « canoniques », comme ce que l'on pourrait trouver en majorité dans des textes d'adultes.

### 3.3.2. Usage des marques graphiques de segmentation

Il arrive fréquemment que les participants utilisent la ponctuation de façon non standard lorsqu'ils segmentent leur texte. Dans ce cadre, nous avons identifié deux types de problème. Premièrement, il peut y avoir un problème au niveau du marquage des PG, avec soit une absence de majuscule (3) soit une absence de point (4). En effet, la présence d'une seule des deux marques était suffisante pour identifier une PG, d'où la possible absence de la seconde marque.

(3) *Je me disais que mon père m'aimerait encore moins. ça m'a fait un sacré choc*. [S34, CM2, narratif]

(4) *J'étais assis sur un banc alors que je parlais avec mon copain, une série de coups de pied me frappa Tous étaient mes amis sauf un*. [S46, 5<sup>e</sup>, narratif]

Deuxièmement, nous avons identifié les problèmes de marquage entre deux PS : celles-ci ont été segmentées en fonction de critères

7. Les erreurs d'orthographe n'ont pas été conservées.

8. Les exemples sont identifiés avec les éléments suivants : [n° du sujet, niveau scolaire, type de texte].



syntactiques, mais il arrive qu'elles ne soient séparées ni par une virgule ni par un connecteur, ce qui n'est pas canonique.

(5) *Il commençait à faire nuît c'était bientôt l'heure de manger.* [S34, 3<sup>e</sup>, narratif]

### 3.3.3. Longueur des pauses

En regard du codage des marques de segmentation ou de leur absence, nous avons examiné la longueur des pauses associées. Pour cela, nous avons pris en compte la pause entre le dernier mot de l'unité et le premier mot de l'unité suivante. Par exemple, dans le cas de deux PG séparées par un point, la « pause entre PG » est constituée de la pause avant le point additionnée de la pause après le point ; si le point était omis entre les deux PG, il s'agissait simplement de la pause entre les deux mots.

Dans un premier temps, nous avons identifié les pauses les plus longues pour chaque texte. Plutôt que de déterminer un seuil valable pour tous les sujets, nous avons déterminé des seuils pour chaque texte. Pour chaque texte, nous avons classé toutes les pauses (c'est-à-dire le temps écoulé entre deux mots) dans l'ordre décroissant. Le seuil pour les pauses les plus longues que nous avons retenu correspond à la durée au-delà de laquelle il y a les 20 % de pauses les plus longues. Ce seuil nous a permis de distinguer les pauses « longues » du texte — faisant partie des 20 % de pauses les plus longues — et les pauses « courtes » — faisant partie des 80 % de pauses les moins longues du texte.

Dans un second temps, nous avons déterminé, pour chaque emplacement syntaxique, dans quelle catégorie de pause se situe le mot : par exemple, parmi toutes les pauses en frontière de PG dans un texte, quel est le pourcentage de pauses qui sont longues. Ainsi, nous avons trois variables exprimées en pourcentages : « frontières de PG – longues », « frontières de PS – longues », « frontières de C – longues ».

## 3.4. Analyses utilisées et présentation des résultats

Nous avons tout d'abord effectué des ANOVA, avec pour variable indépendante le niveau scolaire (3 modalités : CM2, 5<sup>e</sup>, 3<sup>e</sup>). Dans le cas d'un effet significatif du niveau scolaire, des analyses *post hoc* ont permis de déterminer si la différence était significative entre chaque niveau scolaire.

Nous avons ensuite établi des « profils », grâce aux analyses « en nuées dynamiques » (ou *k-means clusters*), en utilisant le logiciel SPSS.

Pour cela, nous avons à chaque fois fait des essais avec 2, 3 et 4 groupes, puis nous avons conservé le nombre de groupes qui paraissait le plus pertinent, au vu du nombre d'individus par groupe et des caractéristiques des profils ainsi déterminés. Les graphiques proposés ont été effectués à partir des centres de classe finaux.

## 4. Résultats

### 4.1. Caractéristiques générales

#### 4.1.1. La segmentation

Nous avons comptabilisé le nombre moyen de mots, de PS et de PG par texte. Le tableau 1 indique les moyennes et les écarts-types de chaque variable en fonction du niveau scolaire.

Il existe un effet du niveau scolaire pour le nombre de mots par texte ( $F(2)=30,60$ ;  $p<0,001$ ). La différence est significative entre chaque niveau scolaire : les élèves de 3<sup>e</sup> écrivent plus de mots que ceux de 5<sup>e</sup>, qui eux-mêmes en écrivent davantage que ceux de CM2.

Concernant le nombre de PG par texte, un effet du niveau scolaire ( $F(2)=28,74$ ;  $p<0,001$ ) révèle que les textes des 3<sup>e</sup> contiennent plus de PG que ceux des 5<sup>e</sup>, qui eux-mêmes en contiennent plus que ceux des CM2.

Un effet significatif du niveau scolaire indique que les textes des élèves de 5<sup>e</sup> et 3<sup>e</sup> sont composés de plus de PS que ceux des élèves de CM2 ( $F(2)=21,60$ ;  $p<0,001$ ).

Enfin, le pourcentage de PG contenant 1 ou 2 PS dépend aussi du niveau scolaire ( $F(2)=10,36$ ;  $p<0,001$ ) : ce pourcentage est plus élevé dans les textes de 5<sup>e</sup> et 3<sup>e</sup> que dans ceux de CM2.

	CM2	5 <sup>e</sup>	3 <sup>e</sup>
nombre de mots par texte	64,45 (34,39)	124,48 (60,13)	163,60 (70,57)
nombre de PG par texte	2,25 (1,36)	5,13 (3,21)	6,65 (2,95)
nombre de PS par texte	5,60 (2,98)	10,63 (6,04)	12,10 (4,38)
% de PG contenant 1 ou 2 PS	54,74 % (32,54)	72,70 % (18,77)	77,61 % (16,38)

Tableau 1. – Segmentation des textes en fonction du niveau scolaire.

#### 4.1.2. Seuil des pauses longues

Pour chaque texte un seuil identifiant les 20 % de pauses les plus longues a été déterminé. Le tableau 2 en indique les moyennes et écarts-types en fonction du niveau scolaire.

Un effet du niveau scolaire a été relevé ( $F(2)=7,60$  ;  $p<0,001$ ) : le seuil déterminant les « pauses longues », est significativement plus élevé dans les textes des élèves de CM2 que de 5<sup>e</sup> et 3<sup>e</sup>.

	CM2	5 <sup>e</sup>	3 <sup>e</sup>
seuil pauses longues (en ms)	1 763,23 (905,55)	1 301,43 (671,182)	1 136,55 (629,68)

Tableau 2. – Seuil des 20 % des pauses les plus longues des textes en fonction du niveau scolaire.

### 4.2. Frontières syntaxiques

#### 4.2.1. Usage des marques graphiques de segmentation

Le tableau 3 indique les moyennes et écarts-types des trois types de problème liés à la ponctuation : absence de virgule et de connecteur à une frontière de PS, absence de majuscule en début de PG ou de point en fin de PG.

Les problèmes de ponctuation en frontière de PS dépendent du niveau scolaire ( $F(2)=9,06$  ;  $p<0,001$ ) : il y a une différence significative du pourcentage d'erreurs entre les textes des CM2 et ceux des 5<sup>e</sup> et 3<sup>e</sup>, les plus jeunes faisant davantage d'erreurs.

Concernant les problèmes de majuscule, il y a un effet du niveau scolaire ( $F(2)=6,65$  ;  $p<0,005$ ) : les élèves de CM2 font davantage d'erreurs que ceux de 5<sup>e</sup> et 3<sup>e</sup>.

Enfin, il y a également un effet du niveau scolaire pour l'absence de point en fin de PG ( $F(2)=7,94$  ;  $p<0,001$ ), les élèves de CM2 faisant davantage ce type d'écart que ceux de 5<sup>e</sup> et de 3<sup>e</sup>.

	CM2	5 <sup>e</sup>	3 <sup>e</sup>
PS juxtaposées sans virgule ni connecteur (% par texte)	24,68 % (21,75)	14,87 % (19,54)	7,04 % (13,44)
PG sans majuscule au début (% par texte)	21,46 % (26,48)	10,22 % (15,60)	5,86 % (15,00)
PG sans point à la fin (% par texte)	12,47 % (18,97)	1,01 % (4,28)	5,05 % (11,50)

Tableau 3. – Problèmes liés à la ponctuation en fonction du niveau scolaire.

Nous avons voulu déterminer des profils de scripteurs en fonction des problèmes de ponctuation. Les analyses menées ont permis d'identifier 3 profils (A1, A2, A3). La figure 1 illustre les caractéristiques des problèmes de ponctuation pour chaque profil établi.

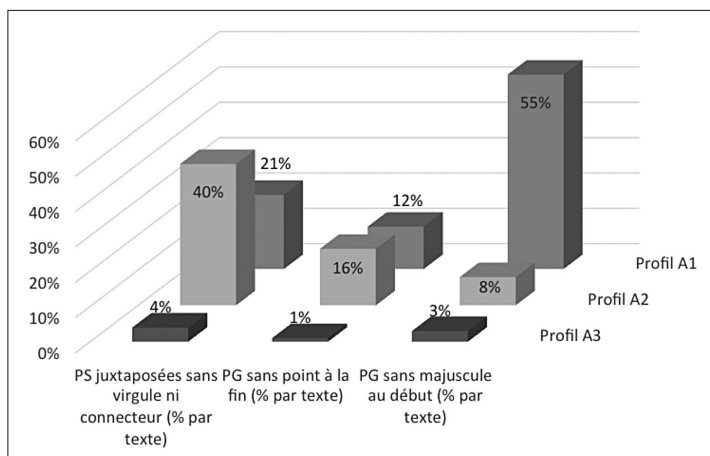


Figure 1. – Profils de scripteurs liés aux problèmes de ponctuation.

Le profil A3 correspond à l'écriture la plus canonique, puisqu'il y a peu de problèmes de ponctuation. Le profil A2 correspond aux élèves faisant essentiellement des erreurs de virgules (absence de virgules et de connecteurs en frontière de PS). Enfin, les élèves du profil A1 sont ceux qui s'écartent le plus de la norme d'un point de vue quantitatif, leurs difficultés principales se situant au niveau de l'emploi des majuscules en début de PG.

Le tableau 4 indique le nombre de scripteurs par profil en fonction du niveau scolaire : par exemple, très peu d'élèves de 3<sup>e</sup> ont le profil A1 ou A2, ces deux profils étant surtout répandus parmi les élèves de CM2.

	A1 : beaucoup de problèmes de majuscules	A2 : beaucoup de problèmes de virgules	A3 : peu de problèmes de ponctuation
CM2	13	17	10
5 <sup>e</sup>	4	8	28
3 <sup>e</sup>	2	4	34
Total	19	29	72

Tableau 4. – Profils de scripteurs liés aux problèmes de ponctuation en fonction du niveau scolaire.

#### 4.2.2. Longueur des pauses suivant l'emplacement syntaxique

Pour chaque emplacement syntaxique (frontière de PG, de PS ou de C), nous avons relevé la proportion de pauses longues. Les moyennes et écarts-types sont donnés dans le tableau 5.

Il y a un effet du niveau scolaire ( $F(2)=40,64$ ;  $p<0,001$ ) pour la proportion de frontières de PG – longues, avec une augmentation de cette proportion en fonction du niveau scolaire, la différence étant significative entre chaque groupe.

En revanche, il n'y a aucun effet significatif pour la proportion de frontières de PS – longues.

Concernant la proportion de frontières de C – longues, un effet du niveau scolaire a été relevé ( $F(2)=2,94$ ;  $p=0,05$ ) : cette proportion est plus élevée chez les CM2 que chez les élèves de 5<sup>e</sup> et de 3<sup>e</sup>.

	CM2	5 <sup>e</sup>	3 <sup>e</sup>
Frontière de PG – longues	40,36 % (26,70 %)	65,91 % (17,38 %)	78,13 % (9,02 %)
Frontière de PS – longues	56,16 % (28,92 %)	56,26 % (20,28 %)	58,17 % (20,11 %)
Frontière de C – longues	41,85 % (24,43 %)	32,25 % (15,42 %)	35,50 % (11,79 %)

Tableau 5. – Longueur des pauses suivant l'emplacement syntaxique en fonction du niveau scolaire.

Nous avons de nouveau établi des profils de scripteurs, cette fois en fonction de la longueur des pauses aux différents emplacements syntaxiques. La figure 2 illustre les caractéristiques des longueurs des pauses des scripteurs des trois profils obtenus (B1, B2, B3).

Le profil B3 correspond à des scripteurs marquant beaucoup de pauses longues entre les PG (65 %), et davantage de pauses courtes entre les PS et les C. Ce profil est l'inverse du profil B1, dans lequel les scripteurs marquent peu de pauses longues entre les PG (22 %), mais beaucoup de pauses longues entre les PS et un nombre assez élevé de pauses longues entre C. Les scripteurs du profil B2, quant à eux, font beaucoup de pauses longues entre les PG et les PS (respectivement 72 % et 67 %), et plus de pauses courtes entre les C.

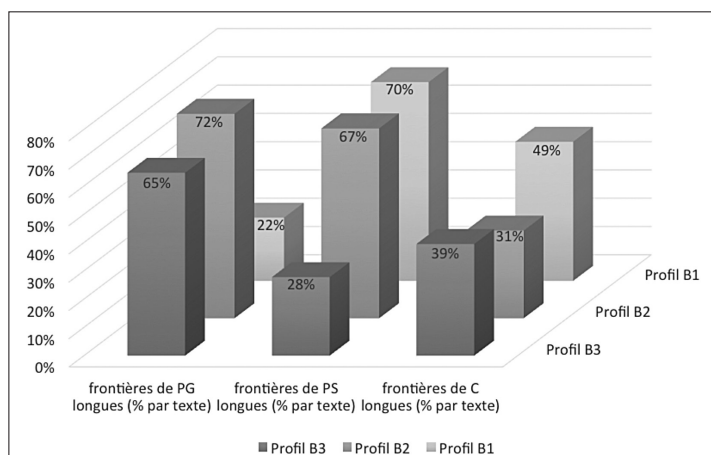


Figure 2. – Profils de scripteurs liés à la longueur des pauses suivant l'emplacement syntaxique.

Le tableau 6 indique le nombre de scripteurs par profil en fonction du niveau scolaire : par exemple, aucun élève de 3<sup>e</sup> n'a le profil B1, ce profil étant surtout trouvé chez les élèves de CM2.

	B1 : frontières de PG courtes, frontières de PS longues	B2 : frontières de PG et de PS longues	B3 : frontières de PG longues
CM2	17	10	13
5 <sup>e</sup>	4	27	9
3 <sup>e</sup>	0	29	11
Total	21	66	33

Tableau 6. – Profils de scripteurs liés la longueur des pauses suivant l'emplacement syntaxique en fonction du niveau scolaire.

#### 4.3. Lien entre produit final et gestion chronométrique

Nous avons enfin souhaité voir si, de manière générale, les scripteurs ayant des problèmes d'utilisation de la ponctuation étaient ceux qui avaient une gestion des pauses moins canonique (avec peu de pauses longues entre PG). Nous avons donc établi le tableau 7 qui croise les deux profils, en indiquant le nombre d'élèves par niveau scolaire pour chaque case. Par exemple, 21 élèves ont peu de problèmes de ponctuation (profil A3) et des frontières de PG longues (profil B3), dont 5 sont en CM2, 6 sont en 5<sup>e</sup> et 10 sont en 3<sup>e</sup>.

	A1 : beaucoup de problèmes de majuscule	A2 : beaucoup de problèmes de virgule	A3 : peu de problèmes de ponctuation	Total
B1 : frontières de PG courtes, frontières de PS longues	7 (7 en CM2)	8 (7 en CM2, 1 en 5 <sup>e</sup> )	6 (3 en CM2, 3 en 5 <sup>e</sup> )	21
B2 : frontières de PG et PS longues	8 (3 en CM2, 3 en 5 <sup>e</sup> , 2 en 3 <sup>e</sup> )	13 (5 en CM2, 5 en 5 <sup>e</sup> , 3 en 3 <sup>e</sup> )	45 (2 en CM2, 19 en 5 <sup>e</sup> , 21 en 3 <sup>e</sup> )	66
B3 : frontières de PG longues	4 (3 en CM2, 1 en 3 <sup>e</sup> )	8 (5 en CM2, 2 en 5 <sup>e</sup> , 1 en 3 <sup>e</sup> )	21 (5 en CM2, 6 en 5 <sup>e</sup> , 10 en 3 <sup>e</sup> )	33
Total	19	29	72	120

Tableau 7. – Croisement des profils liés à la ponctuation et à la longueur des pauses.

## 5. Discussion

Nous nous sommes attachés ici à la segmentation de textes d'élèves telle qu'elle se manifeste à travers la ponctuation et l'organisation syntaxique des clauses en tentant de la mettre en lien avec des mesures chronométriques de l'activité d'écriture. Dans ce travail, nous avons d'abord établi que les textes s'allongent avec les années de scolarisation (niveau scolaire). Les textes des élèves de CM2 sont en effet en moyenne deux fois moins longs (en nombre de mots) que les textes des élèves de 5<sup>e</sup> et les textes des élèves de 3<sup>e</sup> sont en moyenne 30 % plus longs que ceux de 5<sup>e</sup>. Si les textes sont plus longs, cela implique plus de possibilités de segmentation et l'on n'est pas étonné de constater que les élèves de CM2 ont en moyenne 2 fois moins de PG que les 5<sup>e</sup> et les 3<sup>e</sup> ont environ 30 % de PG de plus que les 5<sup>e</sup>. Les proportions ne sont pas si bien conservées pour les PS puisque si les CM2 ont bien environ deux fois moins de PS par texte que les 5<sup>e</sup>, les textes des élèves de 3<sup>e</sup> comportent en moyenne seulement environ 14 % de plus de PS que les textes de 5<sup>e</sup> ; il conviendrait donc d'examiner plus particulièrement la longueur des PG et des PS à travers les textes. Enfin, la proportion de PG contenant une ou deux PS, que nous avons identifiée comme un indice de conformité à la norme de segmentation, augmente selon le niveau scolaire, surtout entre les élèves de CM2 et ceux de 5<sup>e</sup> et 3<sup>e</sup>.

Concernant les mesures chronométriques, nous avons choisi d'examiner la même proportion de pauses pour chaque texte (les 20 % de

pauses les plus longues), ce qui nous conduit à nous référer à un seuil différent pour chaque texte. Les résultats montrent que pour les 3 groupes le seuil moyen pour les pauses longues est à plus de 1 seconde, qui est le seuil le plus communément utilisé pour les adultes dans la littérature (Schilperoord, 2002). Il est à noter également que le seuil moyen diminue à mesure que le niveau scolaire augmente. Dans le cadre de la théorie de la capacité (McCutchen, 1996), cette diminution pourrait être mise en relation avec des changements au niveau des coûts cognitifs : par exemple, au fil de l'expérience certaines opérations gagnent en automaticité.

L'usage des marques graphiques de segmentation, nous l'avons vu, présente des écarts par rapport à la norme dans nos textes. Nous avons codé l'absence de virgule et de connecteur à une frontière de PS, l'absence de majuscule en début de PG ou l'absence de point en fin de PG. Les résultats indiquent une baisse moyenne des écarts par rapport à la norme concernant les deux premiers indicateurs (virgules, majuscules) avec le niveau scolaire. L'absence de point est l'écart par rapport à la norme le moins présent dans nos textes et ne fait de différence qu'entre les CM2 et les élèves plus âgés (5<sup>e</sup> et 3<sup>e</sup>). Cela rejoint des études qui ont montré que la ponctuation des phrases était mal maîtrisée en 6<sup>e</sup>, même si les élèves savent, d'un point de vue théorique, qu'une PG est limitée par une majuscule et un point (Fauchart, 2001). Toutefois, l'examen des moyennes masque l'existence de profils de scripteurs selon ces dimensions. Nous avons ainsi repéré des scripteurs qui avaient globalement peu de difficultés visibles dans les textes dont seulement 10 sur 72 étaient en CM2 ; une trentaine de scripteurs ont été identifiés comme ayant des difficultés quant aux virgules et connecteurs en frontière de PS et enfin une vingtaine d'élèves marquent peu les majuscules en début de PG. Dans ces deux derniers profils se trouvent essentiellement des élèves de CM2.

En ce qui concerne la longueur des pauses en fonction du type de frontière syntaxique, les résultats suggèrent que le travail de planification est mieux reporté sur des frontières fortes en 5<sup>e</sup> et en 3<sup>e</sup> qu'en CM2. Une pause longue suggère en effet un coût cognitif important, pouvant correspondre par exemple à une activité de planification (Foulin & Fayol, 1988 ; Passerault 1991) ou de relecture de ce qui vient d'être écrit. La proportion de frontières de PG correspondant à des pauses longues étant plus importante en 3<sup>e</sup> qu'en 5<sup>e</sup>, et plus en 5<sup>e</sup> qu'en CM2, cela suggère une plus forte adéquation de la planification à des frontières fortes. Les frontières de PS correspondant à des pauses longues



connaissent sensiblement la même répartition à travers les niveaux scolaires, ce qui serait conforme à nos attentes, les frontières de PS étant des frontières fortes. Enfin, les frontières de C marquées par des pauses longues sont en proportion moindre en 3<sup>e</sup> et cela conforterait notre interprétation, les frontières de C étant moins fortes. Les profils de scripteurs qui émergent des données sont d'une part des scripteurs qui planifieraient en frontière de C ou de PS, stratégie fréquente chez les élèves de CM2 ; des scripteurs qui planifieraient en frontière de PS ou PG, stratégie fréquente chez les élèves de 5<sup>e</sup> et de 3<sup>e</sup> ; ou enfin, des scripteurs qui planifieraient essentiellement en frontière de PG. Outre une normalisation de la segmentation (Paolacci & Favart, 2010), les élèves tendent donc vers un profil chronométrique plus proche de celui des adultes, avec des planifications plutôt au niveau des PS voire des PG, ce qui corrobore une étude menée précédemment sur les unités de planification (Ailhaud et coll., 2016). Certes, les longues pauses peuvent parfois servir à relire ce qui a été écrit plutôt qu'à planifier la suite, mais le choix de s'arrêter en frontière d'unité syntaxique implique que cette unité peut être considérée comme une unité de production.

En mettant en lien les profils chronométriques et les profils quant aux écarts par rapport à la norme, on remarque que les élèves qui répondent le mieux à la norme entrent dans des profils qui ont des proportions importantes de pauses longues aux frontières de PG et de PS. Plus particulièrement, un nombre assez important de personnes dont les textes ont une proportion élevée de pauses longues en frontières de C et de PS (profil B1) font beaucoup d'erreurs soit de virgule, soit de majuscule (profils A1 et A2) ; sur ces 15 élèves, 14 sont en CM2. À l'inverse, parmi ceux qui ont peu de problèmes de ponctuation (profil A3), une très grande partie (66 élèves) font des pauses longues en début de PS et de PG (profil B2) ; parmi eux 19 sont en 5<sup>e</sup> et 24 sont en 3<sup>e</sup>. L'interprétation du profil de scripteur faisant des pauses longues en frontières de PG (profil B3) est plus difficile : le nombre d'élèves de ce profil est assez stable en fonction des niveaux. Cependant, alors que 10 élèves sur 11 en 3<sup>e</sup> ont peu de problèmes de ponctuation, les élèves de CM2 ayant ce profil ont majoritairement des problèmes avec les virgules ou les majuscules. Une limite importante pour cette interprétation est le fait que des textes de longueurs différentes composent notre corpus. La prise en compte des phénomènes de façon proportionnelle ne nous garantit en effet pas que leur gestion dans le temps puisse être comparable. Par exemple, Fayol (1989) montre que, dans les narrations, la ponctuation se manifeste aux frontières des épisodes, délimitant ainsi

des parties dans le texte, dont certaines sont « sur-ponctuées » (le résumé initial et la fin du récit) et d'autres ne présentent pas de ponctuation (par exemple, pas de ponctuation de la complication jusqu'au dénouement). La proportion de texte constituant les épisodes pourrait ainsi affecter nos résultats.

## 6. Conclusion

Les analyses menées dans cette étude confirment la pertinence de recourir à la segmentation en « phrases syntaxiques » (Paolacci & Rossi-Gensane, 2014) afin de mieux rendre compte des comportements des scripteurs. En effet, la PS semble être l'unité de planification de la majorité des scripteurs à partir de la classe de 5<sup>e</sup>. Cette unité a en outre l'avantage d'être définie uniquement en termes syntaxiques, contrairement à la « phrase », dont la définition mêle des critères sémantiques, prosodiques, énonciatifs, et dont la segmentation prend davantage de sens lorsqu'elle est considérée au sein du texte, et non « en soi » (Rondelli, 2013). Une fois les PS délimitées, il serait possible de discuter avec les élèves de la meilleure marque de segmentation à mettre : en effet, un enchaînement de PS apparaît toujours comme non canonique s'il n'y a ni marque de ponctuation ni connecteur logique ; or, les problèmes de marquage des frontières PS restent présents même pour certains élèves de 3<sup>e</sup>. En discutant de la meilleure marque à mettre à cette frontière, il ne s'agirait plus d'aborder la ponctuation comme marque de segmentation, mais plutôt comme marque de liaison. En effet, cela pourrait ouvrir la discussion sur la diversité des marques de ponctuation en tant que reflet du degré de liaison entre les unités, par exemple suivant l'organisation hiérarchique du texte (Fayol, Carré & Simon-Thibault, 2014), ce qui permettrait aussi de développer l'usage du point-virgule, souvent rare dans les productions d'élèves. Par ailleurs, les unités sont également liées suivant le sémantisme : dans ce cas, l'usage d'un connecteur approprié peut être davantage informatif qu'une virgule ou un point ; là encore, des marques de ponctuation plus rarement utilisées, comme les deux points, pourraient être abordées à travers cette fonction. Dès lors, l'absence de PG dans un texte devrait poser question à un élève : cela signifierait que son texte manque d'articulation, si tout est uniquement relié par des virgules ou la conjonction « et ». Un travail d'oralisation pourrait alors permettre d'identifier où, spontanément, l'enfant marquerait une pause un peu plus longue, ce qui correspondrait à une frontière de PG. Peu à peu, la PG pourrait ainsi prendre sens, se rapprochant

de la notion de « période » (en tant qu'unité prosodique), avec en son sein une ou plusieurs PS correctement reliées les unes aux autres. Nous pouvons supposer que ce n'est que plus tard que, la PG ayant pris sens pour le scripteur, ce dernier sera en mesure de planifier davantage au niveau de la PG que de la PS, adoptant un comportement plus proche de celui des adultes.

Cette étude gagnerait à intégrer la question des connecteurs ainsi que celle de la disposition graphique des clauses (par exemple, les regroupements en paragraphes). Enfin, nous disposons pour chaque élève d'un texte narratif et d'un texte expositif qui ont été considérés ensemble dans cette étude et qui pourraient, considérés séparément, apporter un autre éclairage sur la segmentation.

## RÉFÉRENCES BIBLIOGRAPHIQUES

- AILHAUD, Émilie, CHENU, Florence & JISA, Harriet. (2016). A Developmental Perspective on the Units of Written French. Dans J. Perera, M. Aparici, E. Rosado & N. Salas (dir.), *Written and Spoken Language Development across the Lifespan: Essays in Honour of Liliana Tolchinsky* (p. 287-305). Springer.
- BÉGUELIN, Marie-José. (2002). Clause, période ou autre ? La phrase graphique et la question des niveaux d'analyse. *Verbum*, 24(1-2), 85-107.
- BERMAN, Ruth. (2006). Developing Linguistic Knowledge and Language Use across Adolescence. Dans E. Hoff & M. Shatz (dir.), *Handbook of Language Development* (p. 92-111). Blackwell Publishing.
- BERMAN, Ruth & SLOBIN, Dan. (1994). *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Hillsdale, NJ : Lawrence Erlbaum Associates.
- BERMAN, Ruth & RAVID, Dorit. (2009). Becoming a Literate Language User: Oral and Written Text Construction across Adolescence. Dans D. Olson & N. Torrance (dir.), *Cambridge Handbook of Literacy* (p. 92-111). Cambridge University Press.
- BERRENDONNER, Alain. (2002). Les deux syntaxes. *Verbum*, 24(1-2), 23-36.
- CHESNET, David & ALAMARGOT, Denis. (2005). Analyses en temps réel des activités oculaires et graphomotrices du scripteur : intérêt du dispositif « Eye and Pen ». *L'année psychologique*, 105(3), 477-520.
- CREISSELS, Denis. (1995). *Éléments de syntaxe générale*. Paris : Presses universitaires de France.
- CREISSELS, Denis. (2006). *Syntaxe générale – Une introduction typologique, 1 : catégories et constructions (Lavoisier)*. Paris : Hermès.

- DUFOUR, Marie-Pierre. (2014). *Des traités de ponctuation à la classe de français : didactisation d'un objet de savoir* (Mémoire de maîtrise). Université de Laval, Québec.
- FAUCHART, Nicole. (2001). Le travail de la phrase dans une grammaire pour écrire. *Le français aujourd'hui*, 135, 64-71.
- FAYOL, Michel. (1989). Une approche psycholinguistique de la ponctuation. Étude en production et compréhension. *Langue française*, 81, 21-39.
- FAYOL, Michel, CARRÉ, Mariane & SIMON-THIBULT, Laurence. (2013). Enseigner la ponctuation : comment et avec quels effets ? *Le français aujourd'hui*, 187, 31-40.
- FOULIN, Jean-Noël & FAYOL, Michel. (1988). Étude en temps réel de la production écrite chez des enfants de sept et huit ans. *European Journal of Psychology of Education*, 3(4), 461-475.
- GAYRAUD, Frédérique. (2000). *Le développement de la différenciation orale écrit vu à travers le lexique* (Thèse de doctorat). Université Lumière – Lyon 2, Lyon.
- HUNT, Kellogg W. (1970). Syntactic Maturity in Schoolchildren and Adults. *Monographs of the Society for Research in Child Development*, 35(1).
- KLEIBER, Georges. (2003). Faut-il dire adieu à la phrase ? *L'information grammaticale*, 98, 17-22.
- LE GOFFIC, Pierre. (1993). *Grammaire de la phrase française*. Paris : Hachette.
- MAGGIO, Séverine, LÉTÉ, Bernard, CHENU, Florence, JISA, Harriet & FAYOL, Michel. (2012). Tracking the Mind during Writing: Immediacy, Delayed, and Anticipatory Effects on Pauses and Writing Rate. *Reading and Writing*, 25(9), 2131-2151.
- MCCUTCHEN, Deborah. (1996). A Capacity Theory of Writing: Working Memory in Composition. *Educational Psychology Review*, 8(3), 299-325.
- PAOLACCI, Véronique & FAVART, Monik. (2010). Traitement des marques de cohésion par les jeunes scripteurs : l'utilisation de la ponctuation et des connecteurs à l'entrée en sixième. Approche linguistique, cognitive et didactique. *Langages*, 177(1), 113-128.
- PAOLACCI, Véronique & ROSSI-GENSANE, Nathalie. (2012). Quelles images de la phrase dans les écrits d'élèves de fin d'école primaire française ? Description linguistique et réponses didactiques aux difficultés des élèves. *Actes du 3<sup>e</sup> Congrès mondial de linguistique française*, (vol. 1, p. 341-359). EDP Sciences. <<http://dx.doi.org/10.1051/shsconf/20120100143>>.
- PAOLACCI, Véronique & ROSSI-GENSANE, Nathalie. (2014). Ponctuation et écrits d'élèves : une conception différente de la phrase pour enseigner la ponctuation autrement. *Le français aujourd'hui*, 187, 115-125.

- PASSERAULT, Jean-Michel. (1991). La ponctuation : recherches en psychologie du langage. *Pratiques*, 70, 85-104.
- RAVID, Dorit. (2005). Emergence of Linguistic Complexity in Later Language Development: Evidence from Expository Text Construction. Dans D. Ravid & H. B. Shyldkrot (dir.), *Perspectives on Language and Language Development: Essays in Honor of Ruth A. Berman* (p. 337-355). Dordrecht : Kluwer Academic Publishers.
- RAVID, Dorit & TOLCHINSKY, Liliana. (2002). Developing Linguistic Literacy: A Comprehensive Model. *Journal of Child Language*, 29, 417-447.
- RONDELLI, Fabienne. (2013). La phrase, segment textuel de base ? Choix d'écriture d'élèves de cycle 3 et jugements des enseignants. *Le français aujourd'hui*, 181, 71-81.
- SCHILPEROORD, Joost. (2002). On the Cognitive Status of Pauses in Discourse Production. Dans T. Olive & C. Michael Levy (dir.), *Contemporary Tools and Techniques for Studying Writing* (p. 61-87). Dordrecht : Kluwer Academic Publishers.
- TOLCHINSKY, Liliana. (2006). The Emergence of Writing. Dans C. A. MacArthur, S. Graham & J. Fitzgerald (dir.), *Handbook of Writing Research* (p. 83-95). New York, Londres : The Guilford Press.
- VIGUÉ-SIMON, Anne. (2001). *Développement de la connectivité selon le type textuel et le mode de production au cours de l'adolescence* (Thèse de doctorat). Université Lumière – Lyon 2, Lyon.

## ANNEXE

### Exemple de texte [S13, 5<sup>e</sup>, expositif]

[PG] Les gens ne doivent pas faire la guerre [PS] car ils se tuent entre eux. [PG] Dans les collèges, lycées, écoles il ne faut pas [C] copier sur le voisin [PS] car après on a des problèmes dans l'école [PS] on a copié sur le voisin [PS] donc on n'apprend rien. [PG] Les élèves pourraient éviter de se bagarrer pour des histoires [PS] car quand on est grand après [C] on se tue avec des armes.

[PG] : frontière de phrase graphique ; [PS] : frontière de phrase syntaxique ; [C] : frontière de clause.

Frontière en gras : pause longue (parmi les 20 % des pauses les plus longues du texte).

Frontière soulignée : problème de ponctuation.